Proceso De Normalización

Bases de datos-Modelo de datos relaccional-Proceso de normalización

El proceso de normalización es un estándar que consiste, básicamente, en un proceso de conversión de las relaciones entre las entidades, evitando:

- La redundancia de los datos: repetición de datos en un sistema.
- Anomalías de actualización: inconsistencias de los datos como resultado de datos redundantes y actualizaciones parciales.
- Anomalías de borrado: pérdidas no intencionadas de datos debido a que se han borrado otros datos.
- Anomalías de inserción: imposibilidad de adicionar datos en la base de datos debido a la ausencia de otros datos.

Tomando como referencia la tabla siguiente:

AUTORES Y LIBROS				
NOMBRE	**NACION**	**CODLIBRO**	**TITULO**	**EDITOR**
Date	USA	999	IBD	AW
Ad.Mig.	ESP	888	CyD	RM
Ma.Piat.	ITA	777	CyD	RM
Date	USA	666	BdD	AW

Se plantean una serie de problemas:

- Redundancia: cuando un autor tiene varios libros, se repite la nacionalidad.
- Anomalías de modificación: Si Ad.Mig. y Ma.Piat. desean cambiar de editor, se modifica en los 2 lugares. A priori no podemos saber cuántos autores tiene un libro. Los errores son frecuentes al olvidar la modificación de un autor. Se pretende modificar en un sólo sitio.
- Anomalías de inserción: Se desea dar de alta un autor sin libros, en un principio. NOMBRE y CODLIBRO son campos clave, una clave no puede tomar valores nulos.

Asegurando:

- Integridad entre los datos: consistencia de la información.

El proceso de normalización nos conduce hasta el modelo físico de datos y consta de varias fases denominadas formas normales, estas formas se detallan a continuación.

Definición De La Clave

Antes de proceder a la normalización de la tabla lo primero que debemos de definir es una clave, esta clave deberá contener un valor único para cada registro (no podrán existir dos valores iguales en toda la tabla) y podrá estar formado por un único campo o por un grupo de campos.

En la tabla de alumnos de un centro de estudios no podemos definir como campo clave el nombre del alumno ya que pueden existir varios alumnos con el mismo nombre. Podríamos considerar la posibilidad de definir como clave los campos nombre y apellidos, pero estamos en la misma situación: podría darse el caso de alumnos que tuvieran los mismo apellidos y el mismo nombre (Juan Fernández Martín).

La solución en este caso es asignar un código de alumno a cada uno, un número que identifique al alumno y que estemos seguros que es único.

Una vez definida la clave podremos pasar a estudiar la primera forma normal.

Primera Forma Normal (1NF)

Se dice que una tabla se encuentra en primera forma normal (1NF) si y solo si cada uno de los campos contiene un único valor para un registro determinado. Supongamos que deseamos realizar una tabla para guardar los cursos que están realizando los alumnos de un determinado centro de estudios, podríamos considerar el siguiente diseño:

Código	Nombre	Cursos
1	Marcos	Inglés
2	Lucas	Contabilidad, Informática
3	Marta	Inglés, Contabilidad

Podemos observar que el registro de código 1 si cumple la primera forma normal, cada campo del registro contiene un único dato, pero no ocurre así con los registros 2 y 3 ya que en el campo cursos contiene más de un dato cada uno. La solución en este caso es crear dos tablas del siguiente modo:

TABLA A		TABLA B	
Código	Nombre	Código	Curso
1	Marcos	1	Inglés
2	Lucas	2	Contabilidad
3	Marta	2	Informática
		3	Inglés
		3	Informática

Como se puede comprobar ahora todos los registros de ambas tablas contienen valores únicos en sus campos, por lo tanto ambas tablas cumplen la primera forma normal.

Una vez normalizada la tabla en 1NF, podemos pasar a la segunda forma normal.

Segunda Forma Normal (2NF)

La segunda forma normal compara todos y cada uno de los campos de la tabla con la clave definida. Si todos los campos dependen directamente de la clave se dice que la tabla está es segunda forma normal (2NF).

Supongamos que construimos una tabla con los años que cada empleado ha estado trabajando en cada departamento de una empresa:

Código Empleado	Código Dpto.	Nombre	Departamento	Años
1	6	Juan	Contabilidad	6
2	3	Pedro	Sistemas	3
3	2	Sonia	I+D	1
4	3	Verónica	Sistemas	10
2	6	Pedro	Contabilidad	5

Tomando como punto de partida que la clave de esta tabla está formada por los campos código de empleado y código de departamento, podemos decir que la tabla se encuentra en primera forma normal, por tanto vamos a estudiar la segunda:

1. El campo nombre no depende funcionalmente de toda la clave, sólo depende del código del empleado.
2. El campo departamento no depende funcionalmente de toda la clave, sólo del código del departamento.

3. El campo años si que depende funcionalmente de la clave ya que depende del código del empleado y del código del departamento (representa el número de años que cada empleado ha trabajado en cada departamento)

Por tanto, al no depender todos los campos de la totalidad de la clave la tabla no está en segunda forma normal, la solución es la siguiente:

Tabla A		Tabla B		Tabla C		
Código Empleado	Nombre	Código Departamento	Dpto.	Código Empleado	Código Departamento	Años
1	Juan	2	I+D	1	6	6
2	Pedro	3	Sistemas	2	3	3
3	Sonia	6	Contabilidad	3	2	1
4	Verónica			4	3	10
				2	6	5

Podemos observar que ahora si se encuentras las tres tabla en segunda forma normal, considerando que la tabla A tiene como índice el campo Código Empleado, la tabla B Código Departamento y la tabla C una clave compuesta por los campos Código Empleado y Código Departamento.

Tercera Forma Normal (3NF)

Se dice que una tabla está en tercera forma normal si y solo si los campos de la tabla dependen únicamente de la clave, dicho en otras palabras los campos de las tablas no dependen unos de otros. Tomando como referencia el ejemplo anterior, supongamos que cada alumno sólo puede realizar un único curso a la vez y que deseamos guardar en que aula se imparte el curso. A voz de pronto podemos plantear la siguiente estructura:

Código	Nombre	Curso	Aula
1	Marcos	Informática	Aula A
2	Lucas	Inglés	Aula B
3	Marta	Contabilidad	Aula C

Estudiemos la dependencia de cada campo con respecto a la clave código:

- Nombre depende directamente del código del alumno.
- Curso depende de igual modo del código del alumno.
- El aula, aunque en parte también depende del alumno, está mas ligado al curso que el alumno está realizando.

Por esta última razón se dice que la tabla no está en 3NF. La solución sería la siguiente:

Tabla A			Tabla B	
Código	Nombre	Curso	Curso	Aula
1	Marcos	Informática	Informática	Aula A
2	Lucas	Inglés	Inglés	Aula B
3	Marta	Contabilidad	Contabilidad	Aula C

Una vez conseguida la segunda forma normal, se puede estudiar la cuarta forma normal.

Cuarta Forma Normal (4NF)

Una tabla está en cuarta forma normal si y sólo si para cualquier combinación clave - campo no existen valores duplicados. Veámoslo con un ejemplo:

Geometría		
Figura	**Color**	**Tamaño**
Cuadrado	Rojo	Grande
Cuadrado	Azul	Grande
Cuadrado	Azul	Mediano
Círculo	Blanco	Mediano
Círculo	Azul	Pequeño
Círculo	Azul	Mediano

Comparemos ahora la clave (Figura) con el atributo Tamaño, podemos observar que Cuadrado Grande está repetido; igual pasa con Círculo Azul, entre otras. Estas repeticiones son las que se deben evitar para tener una tabla en 4NF.

La solución en este caso sería la siguiente:

Tamaño		Color	
Figura	**Tamaño**	**Figura**	**Color**
Cuadrado	Grande	Cuadrado	Rojo
Cuadrado	Pequeño	Cuadrado	Azul
Círculo	Mediano	Círculo	Blanco
Círculo	Pequeño	Círculo	Azul

Ahora si tenemos nuestra base de datos en 4NF.

Otras Formas Normales

Existen otras dos formas normales, la llamada quinta forma normal (5FN) que no detallo por su dudoso valor práctico ya que conduce a una gran división de tablas y la forma normal dominio / clave (FNDLL) de la que no existe método alguno para su implantación